

What Is Claimed Is:

- 860627-9062260
1. A method for extracting information from a file, comprising:
processing a file with a printer driver module to generate a representation having a modified format; and
extracting information from said representation.
2. The method of claim 1, wherein said extracted information is a text string.
3. The method of claim 2, wherein said extracted information includes at least one text characteristic of said text string.
4. The method of claim 3, further comprising:
storing said text string and said at least one text characteristic of said text string in a data base.
5. The method of claim 4, wherein processing further comprises:
launching an owning application used to create said file; and
instructing said owning application to print said file using said printer driver module.

6. The method of claim 5, further comprising:

obtaining a current list of files and a previous list of files;

comparing said current list with said previous list to identify new files and modified files; and

repeating the processing, extracting, and storing steps for each new file and each modified file.

7. The method of claim 6, further comprising searching said database for at least one document which includes desired text.

503
C2 8. A set of instructions residing in a storage medium, said set of instructions capable of being executed by a processor to implement a method for extracting information from a file, the method comprising:

processing a file with a printer driver module to generate a modified file having a modified format; and

extracting information from said modified file.

9. The method of claim 8, wherein said extracted information is a text string.

10. The method of claim 9, wherein said extracted information includes at least one text characteristic of said text string.

11. The method of claim 10, further comprising:

storing said text string and said at least one text characteristic of said text string in a data base.

12. The method of claim 11, wherein the processing step further comprises:

launching an owning application used to create said file; and
instructing said owning application to print said file using said printer driver module.

13. The method of claim 12, further comprising:

obtaining a current list of files and a previous list of files;
comparing said current list with said previous list to identify new files and modified files; and
repeating the processing, extracting, and storing steps for each new file and each modified file.

14. The method of claim 13, further comprising searching said database for at least one document which includes desired text.

SUB
A17

15. A system for extracting information from a file, comprising:

a cataloging device, said cataloging device including a processor adapted to be connected to a computer readable memory;

an extraction application module included within said computer readable memory;
an extraction printer driver module included within said computer readable
memory; and
a database adapted to be connected to said cataloging device.

16. The system of claim 15, further comprising:

a network connected to said cataloging device.

17. The system of claim 15, further comprising:

at least one application module included within said computer readable memory.

507
A27 18. An apparatus for extracting information from a file, comprising:

a processor;

a computer readable memory connected to said processor;

an extraction application module included within said computer readable memory;

and

an extraction printer driver module included within said computer readable
memory.

19. The apparatus of claim 18, further comprising:

at least one application module included within said computer readable memory.